

<epam>

Data. DB. DBMS

DB Basics



TRAINING
CENTER

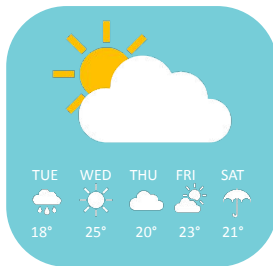
— <epam> —

What is Data?

Minsk, Ivan, 7, Kuprevicha, 80221234567, 220022, Ivanov



0110011101001111110001100111110101110011



5401 0610 0570 0150

What is Information?

Data

Data is a set of facts in a “raw”, non-processed state. Data requires processing to become useful for people.

Minsk, Ivan, 7, Kuprevicha, 80221234567, 220022, Ivanov

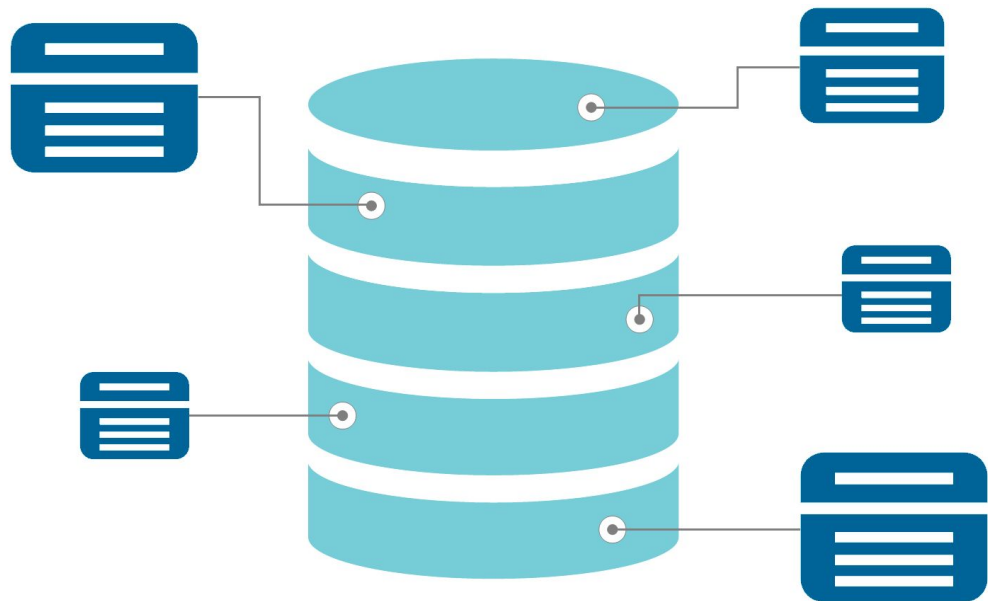
processing

Information

When data is processed, organized, structured and can be used by a human or a PC to gain value / profit.

Ivan Ivanov
Kuprevicha, 7
Minsk, 220022
8-(022)-1234567

What is DataBase (DB)?



Real life examples:

Booking tickets

Weather prediction

Supermarket purchases

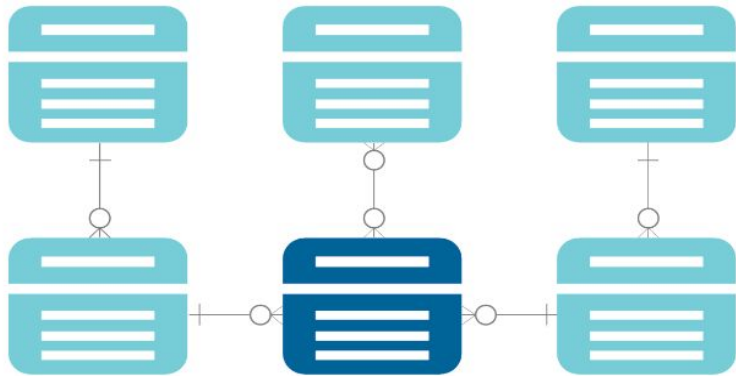
Financial analytics

Instagram

Training courses

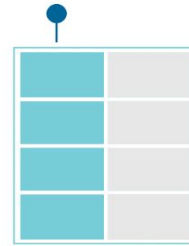
Common DataBase Types

Relational DB

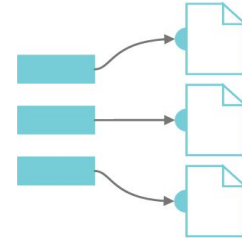


Relational Database Management Systems (RDBMS)

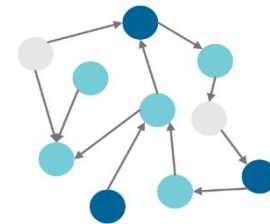
Non-Relational DB



Key-Value



Document



Graph

Common DataBase Types

Trainees				
Name				
Date of Birth				
Gender				
Education				
Prog. experience				

Trainees				
Name	Date of Birth	Gender	Education	Prog. experience
Maria Petrova	14.03.1988	F	Completed High	False
Ivan Ivanov	25.05.1979	M	Incomplete High	True
...	...	...	...	...

Diagram illustrating a database table structure and its components:

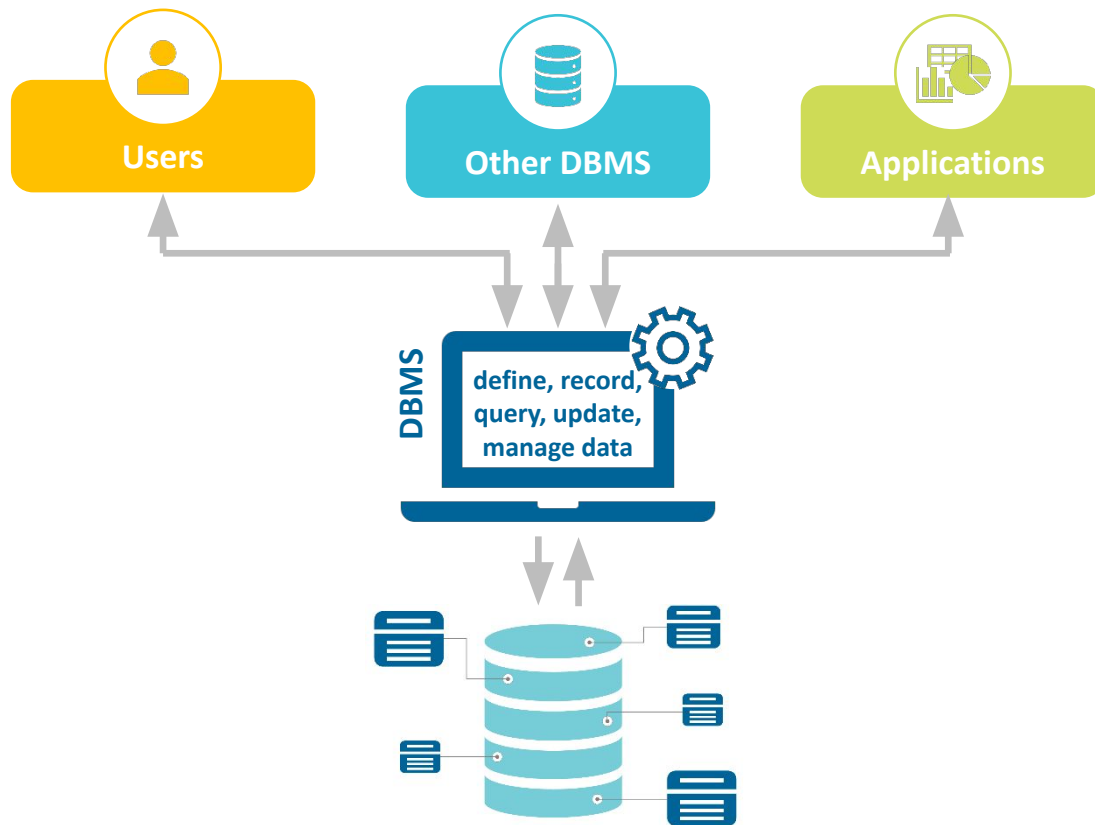
- Table:** The entire structure, labeled "Trainees".
- Columns:** The horizontal headers, labeled "Name", "Date of Birth", "Gender", "Education", and "Prog. experience".
- Rows:** The vertical data entries, labeled "Maria Petrova", "Ivan Ivanov", and "...".
- Cell:** An individual data point, indicated by an arrow pointing to the "True" value in the "Prog. experience" column for "Ivan Ivanov".

Common DataBase Types

Relational database	Key-value database
Table	Collection
Table row	Element of collection
Table column	Key of an element
Table cell	Value by element's key

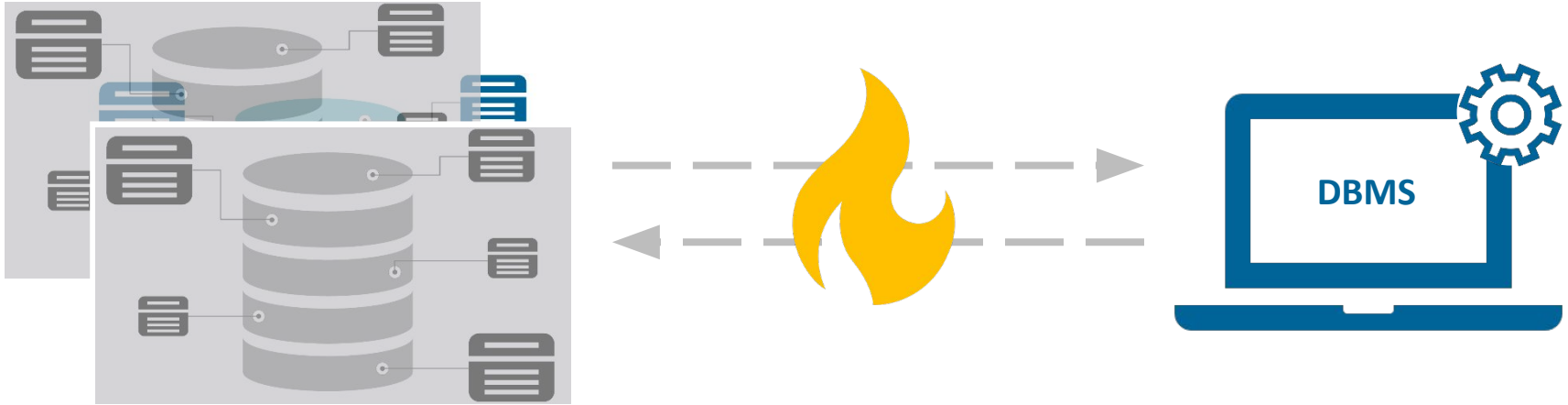
```
"Trainees": [{
  "Name": "Maria Petrova",
  "Date of Birth": "14.03.1988",
  "Gender": "F",
  "Education": "Completed High",
  "Prog. experience": "False"
}, {
  "Name": "Ivan Ivanov",
  "Date of Birth": "25.05.1979",
  "Gender": "M",
  "Education": "Incomplete High",
  "Prog. experience": "True"
}]
```

What is DataBase Management System (DBMS)



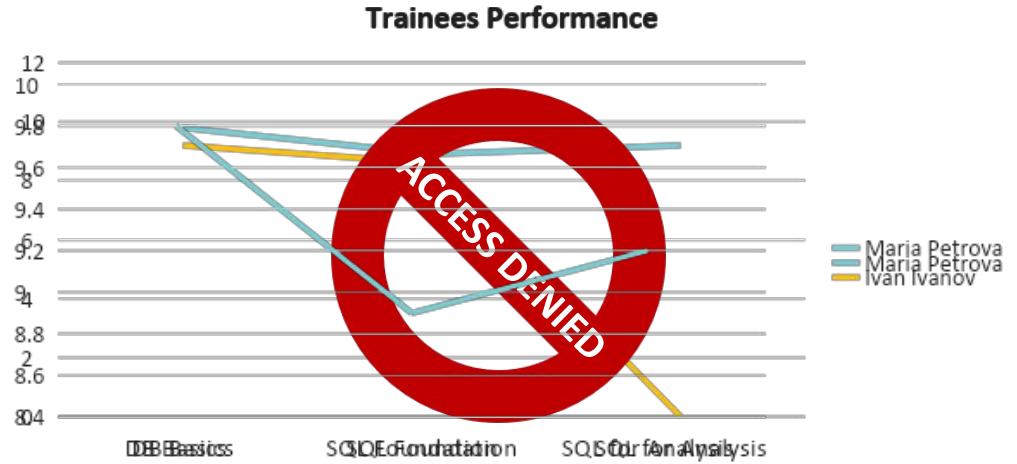
What RDBMS is for?

Manage database backup and recovery processes



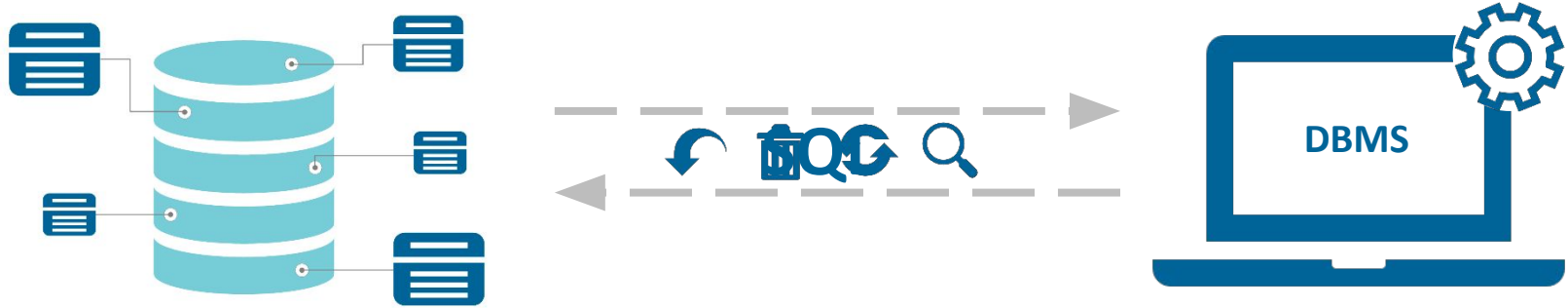
What RDBMS is for?

Restrict data access according to predefined rules



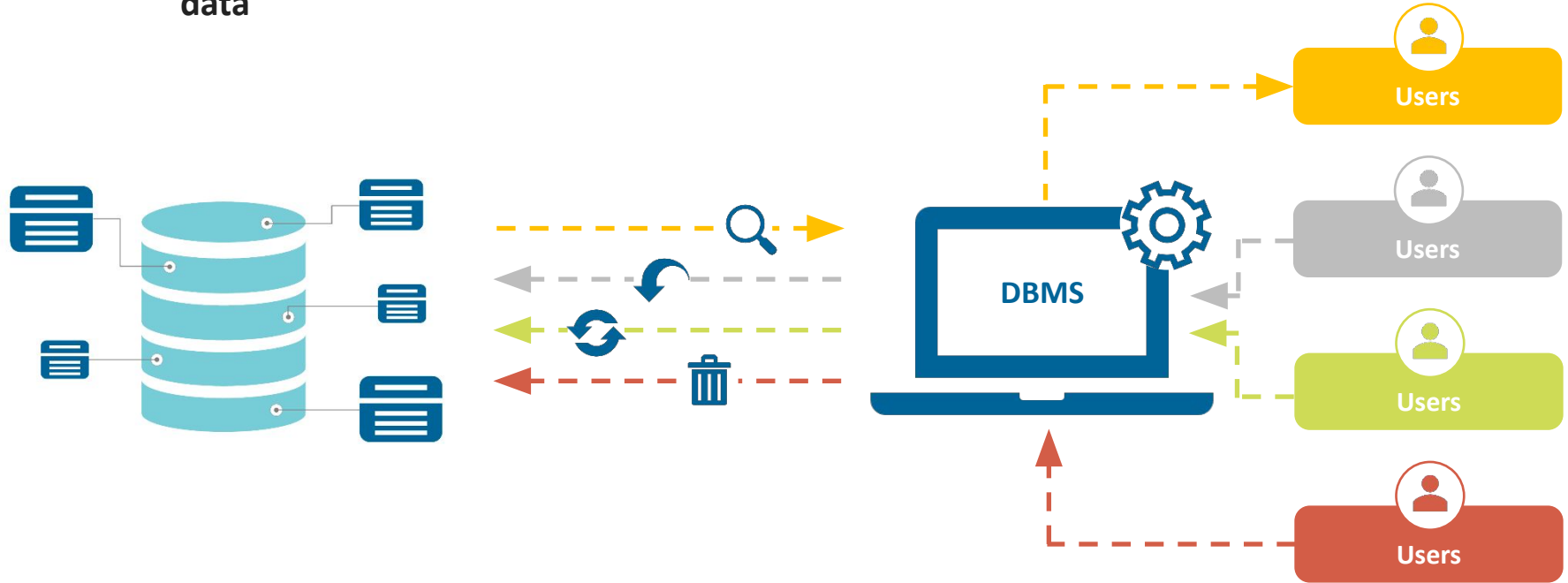
What RDBMS is for?

Allow database access via one of predefined interfaces (most common is SQL query)



What RDBMS is for?

Support data consistency when multiple users work with same pieces of data



<epam>

DB Components

DB Basics



TRAINING
CENTER

— <epam> —

Relational DataBase Components: Table

Name	Date of Birth	Gender	Education	Prog. experience	Hair Color	Favorite Brand
Maria Petrova	14.03.1988	F	Completed High	False	Light brown	Nike
Ivan Ivanov	25.05.1979	M	Incomplete High	True	Blonde	Gant
...	...	...	...	...	...	...

Diagram illustrating data types for the first five columns of the table:

- text**: Points to the **Name** column.
- date**: Points to the **Date of Birth** column.
- boolean**: Points to the **Prog. experience** column.

The last two columns, **Hair Color** and **Favorite Brand**, are highlighted with a blue border.

Relational DataBase Components: Constraint

Name	Date of Birth	Gender	Education	Prog. experience	
Maria Petrova	14.03.1988	F	Completed High	False	😊
Ivan Ivanov	25.05.1979	M	Incomplete High	True	😊
Petr Pirogov	17.11.1987	M	Super High	True	😞
Inna Sidorova	01.01.1992	F	Upper High	False	😞
Volha Hlebava	06.06.1991	F	Completed High	Null	😞
Igor Lesik	15.10.2020	M	Completed High	True	😞

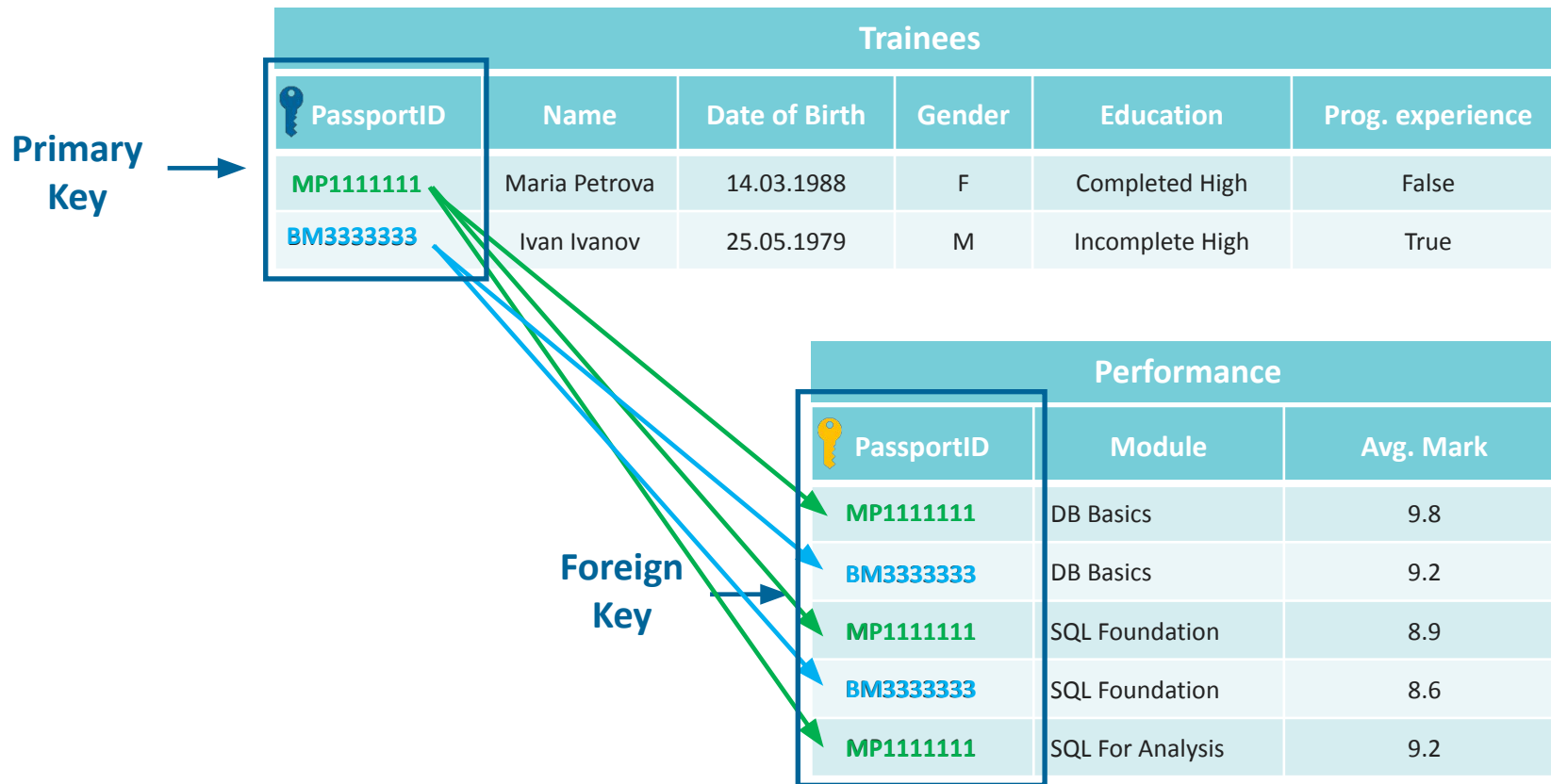
As example: **Current Date = 01.01.2020**

Relational DataBase Components: Keys

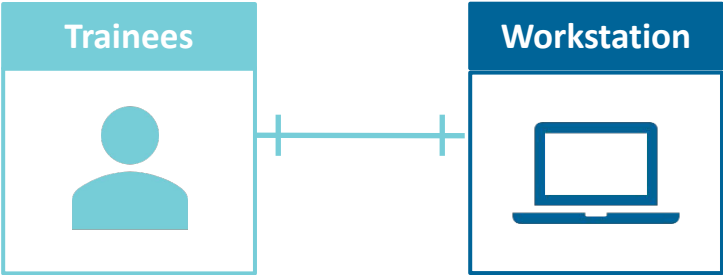
Surrogate Primary Key

PassportID	Name	Date of Birth	Gender	Education	Prog. experience
MP1111111	Maria Petrova	14.03.1988	F	Completed High	False
MP222222	Ivan Ivanov	25.05.1979	M	Incomplete High	True
BM333333	Ivan Ivanov	25.05.1979	M	Incomplete High	True

Relational DataBase Components: Keys



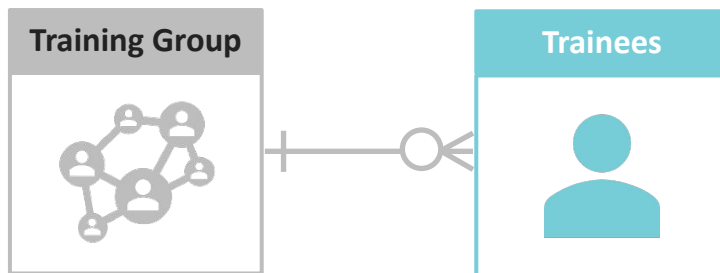
Relational DataBase Components: relationship 1-to-1



Trainees						Workstation	
PassportID	Name	Date of Birth	Gender	Education	Prog. experience	PassportID	Workstation
MP1111111	Maria Petrova	14.03.1988	F	Completed High	False	MP1111111	CCCCCCC4321
BM3333333	Ivan Ivanov	25.05.1979	M	Incomplete High	True	BM3333333	CCCCCCC3322
							CCCCCCC5555



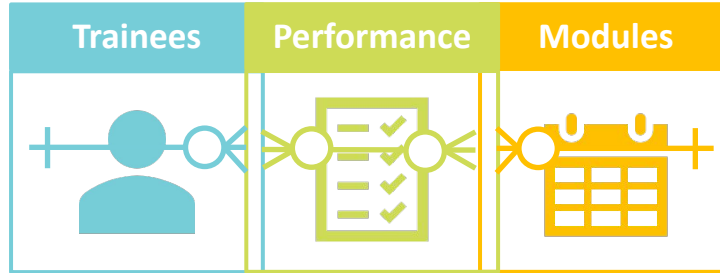
Relational DataBase Components: relationship 1-to-many



Training Group	
GroupID	Group Name
1	Data Training 1
2	.Net Training 1

Trainees						
GroupID	PassportID	Name	Date of Birth	Gender	Education	Prog. exp
1	MP1111111	Maria Petrova	14.03.1988	F	Completed High	False
1	BM3333333	Ivan Ivanov	25.05.1979	M	Incomplete High	True
2	KH8888888	Marta Yakova	96.06.1984	F	Completed High	True

Relational DataBase Components: relationship many-to-many



Trainees		Performance			Modules	
PassportID	Name	PassportID	ModuleID	Avg. Mark	ModuleID	Module Name
MP1111111	Maria Petrova	MP1111111	1	9.8	1	DB Basics
BM3333333	Ivan Ivanov	MP1111111	2	8.9	2	SQL Foundation
		MP1111111	3	9.2	3	SQL for Analysis
		BM3333333	1	9.2		
		BM3333333	2	8.6		

Relational database components

1 Tables

2 Constraints

3 Keys

4 Indexes

5 Views

6 Partitions

<epam>

DB Modeling

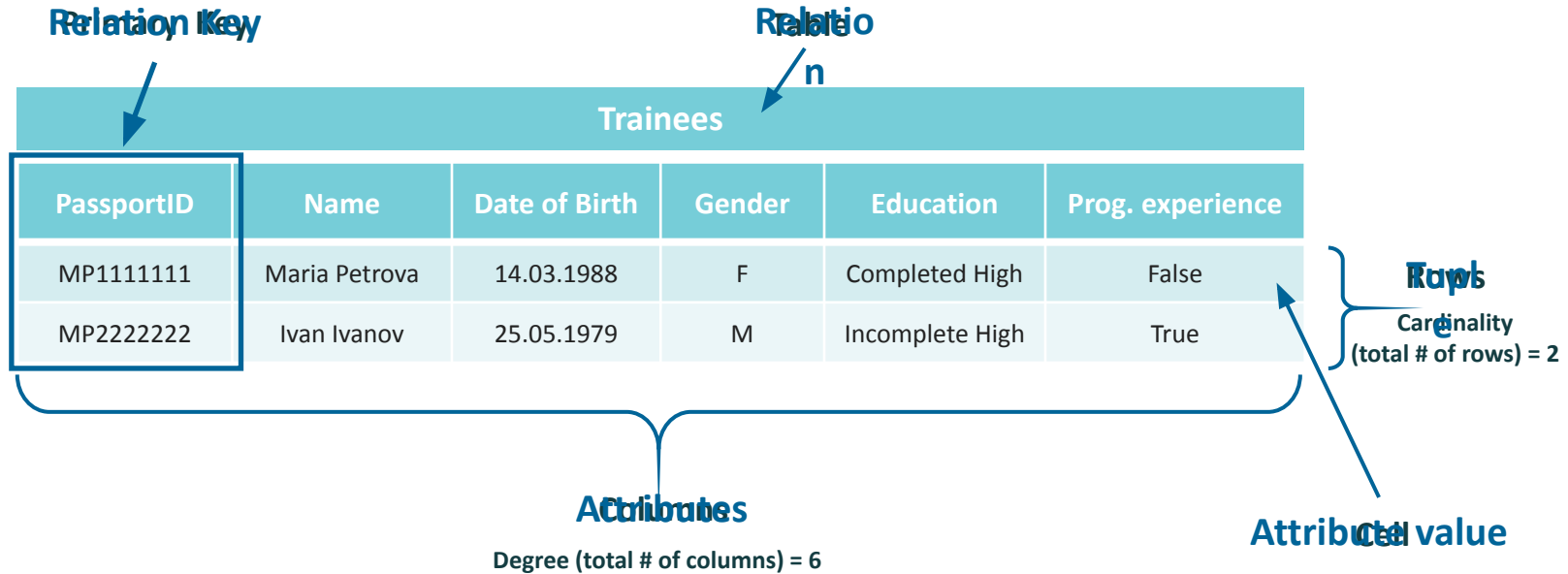
DB Basics



TRAINING
CENTER

— <epam> —

DataBase Structure: Relational Model Terms



DataBase Structure: Rules

Trainees					
PassportID	Name	Date of Birth	Gender	Education	Prog. experience
MP1111111	Maria Petrova	14.03.1988	F	Completed High	False
MP2222222	Ivan Ivanov	25.05.1979	M	Incomplete High	True
MP1313311	Petr Pirogov	17.11.1987	M	Completed High	True
BM1112233	Inna Sidorova	01.01.1992	F	Completed High	False

DataBase Structure: Rules

Trainees					
PassportID	Name	Date of Birth	Gender	Education	Prog. experience
MP1111111	Maria Petrova	14.03.1988	F	Completed High	False
MP2222222	Ivan Ivanov	25.05.1979	M	Incomplete High	True
MP1313311	Petr Pirogov	17.11.1987	M	Completed High	True
BM1112233	Inna Sidorova	01.01.1992	F	Completed High	False
MP2222222	Ivan Ivanov	25.05.1979	M	Incomplete High	True

DataBase Structure: Rules

Trainees					
PassportID	Name	PassportName	Gender	Education	Prog. experience
MP1111111	Maria Petrova	Maria Petrova	F	Completed High	False
MP2222222	Ivan Ivanov	Ivan Ivanov	M	Incomplete High	True
MP1313311	Petr Pirogov	Petr Pirogov	M	Completed High	True
BM1112233	Inna Sidorova	Inna Sidorova	F	Completed High	False

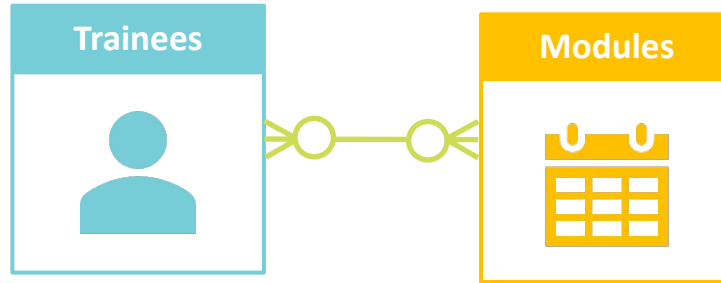
DataBase Structure: Rules

Trainings Participations	
PassportID	Modules
MP1111111	DB Basics, SQL Foundations, SQL for Analysis
MP2222222	DB Basics, SQL Foundations



Trainings Participations	
PassportID	Modules
MP1111111	DB Basics
MP1111111	SQL Foundations
MP1111111	SQL for Analysis
MP2222222	DB Basics
MP2222222	SQL Foundations

DataBase Modeling: Conceptual Model

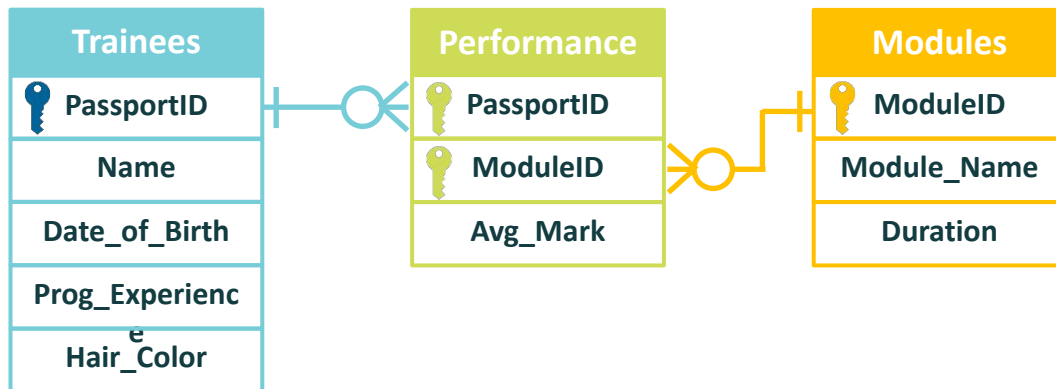


Includes the important entities and the relationships among them

No attributes are specified

No keys are specified

DataBase Modeling: Logical Model



All attributes are specified

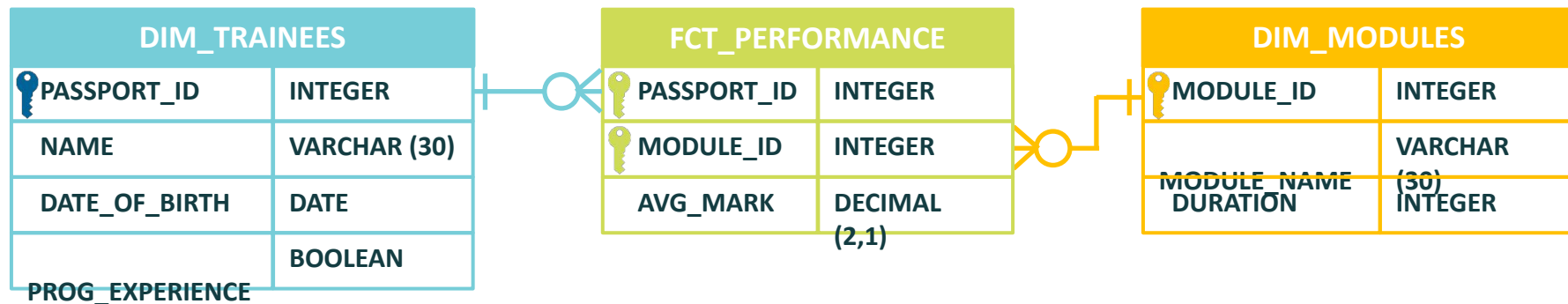
The primary key for each entity specified

Foreign keys are specified

Relationships are specified

Normalization occurs at this level

DataBase Modeling: Physical Model



Convert entities into tables

Convert attributes into columns

Convert relationships into foreign keys

Modify the physical data model based on
physical constraints / requirements

DataBase Modeling: Rules

Trainees					
TrainersID	Name	Date_of_Birth ✓	Gender ✓	Education ✓	! Prog_Experience
MP1111111	Maria Petrova	14.03.1988	F	Completed High	False
MP2222222	Ivan Ivanov	25.05.1979	M	Incomplete High	True
MP1313311	Petr Pirogov	17.11.1987	M	Completed High	True
BM1112233	Inna Sidorova	01.01.1992	F	Completed High	False

~~date~~

Performance		
PassportID	ModuleID	Avg. Mark
MP1111111	1	9.2
MP1111111	2	8.6

~~date~~

~~date~~

DataBase Structure: Rules



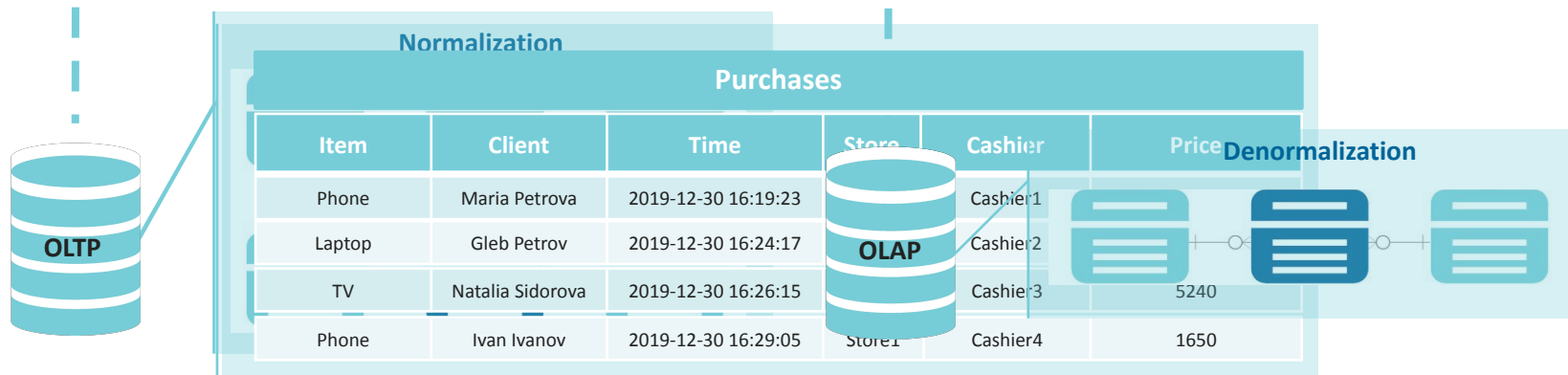
DataBase Tasks



Transactional App



Analytics



<epam>

Normalization

DB Basics



TRAINING
CENTER

— <epam> —

DataBase Modelling Approaches

Trainees					
PassportID	Name	Date_of_Birth	Gender	Education	Prog_Eexperience
MP11111111	Maria Petrovskaya	14.03.1988	F	Completed High	False

Modules	
ModuleID	Module_Name
1	DB Basics
2	SQL Foundation
3	SQL for Analysis

Performance					
PassportID	Name	Date_of_Birth	ModuleID	Module_Name	Avg_Mark
MP11111111	Maria Petrovskaya	14.03.1988	1	DB Basics	9.8
MP11111111	Maria Petrovskaya	14.03.1988	2	SQL Foundation	8.9
MP11111111	Maria Petrovskaya	14.03.1988	3	SQL for Analysis	9.2

Conversion to 1NF

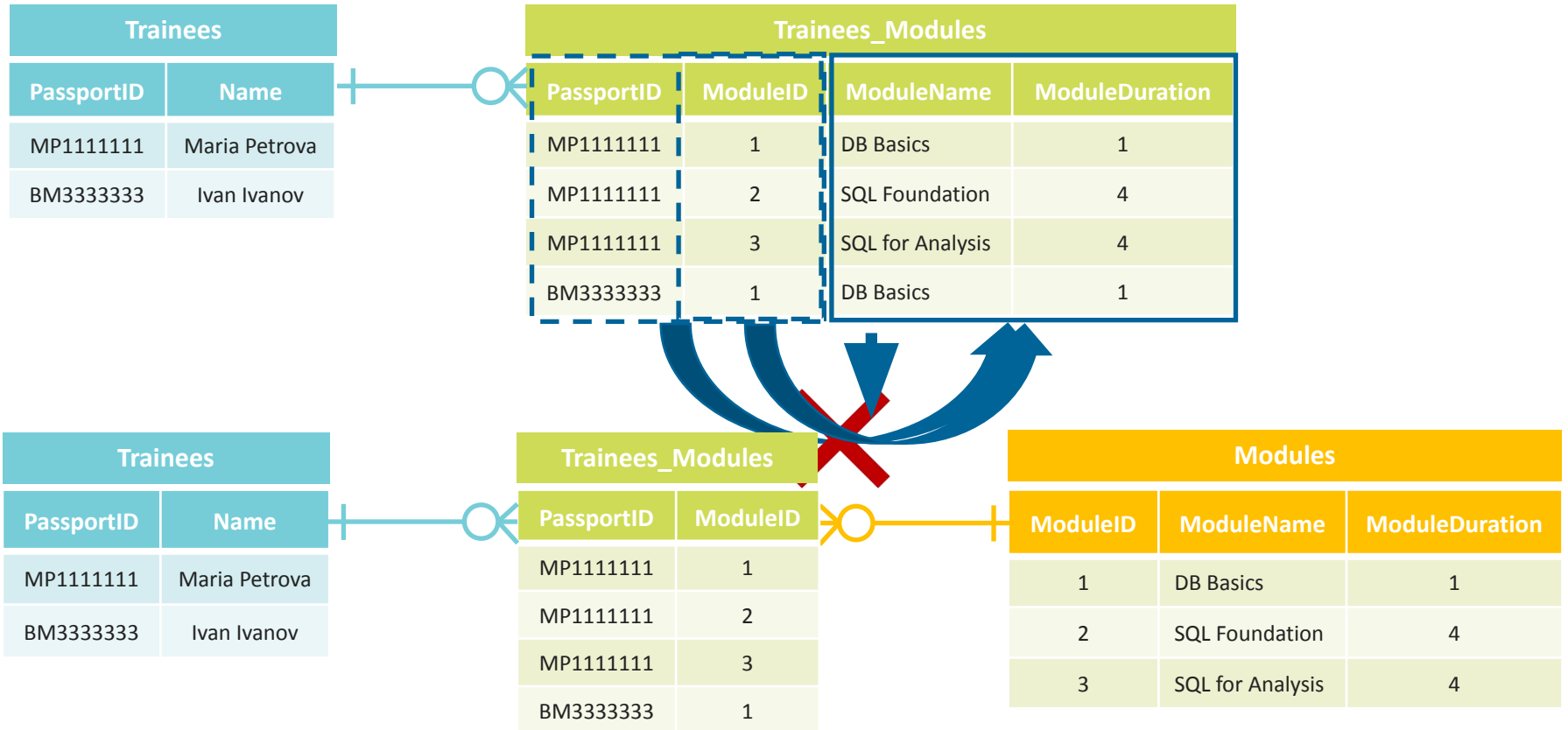
Trainees						
PassportID	Name	Phone_Number	Date_of_Birth	Gender	Education	Prog_experience
MP1111111	Maria Petrova	375291111111	14.03.1988	F	Completed High	False
BM3333333	Ivan Ivanov	375292222222	25.05.1979	M	Incomplete High	True
		375253333333				



Trainees						
PassportID	Name	Phone_Number	Date_of_Birth	Gender	Education	Prog_experience
MP1111111	Maria Petrova	375291111111	14.03.1988	F	Completed High	False
BM3333333	Ivan Ivanov	375292222222	25.05.1979	M	Incomplete High	True
BM3333333	Ivan Ivanov	375253333333	25.05.1979	M	Incomplete High	True

Conversion from 1NF to 2NF

Conversion from 1NF to 2NF



Conversion from 2NF to 3NF

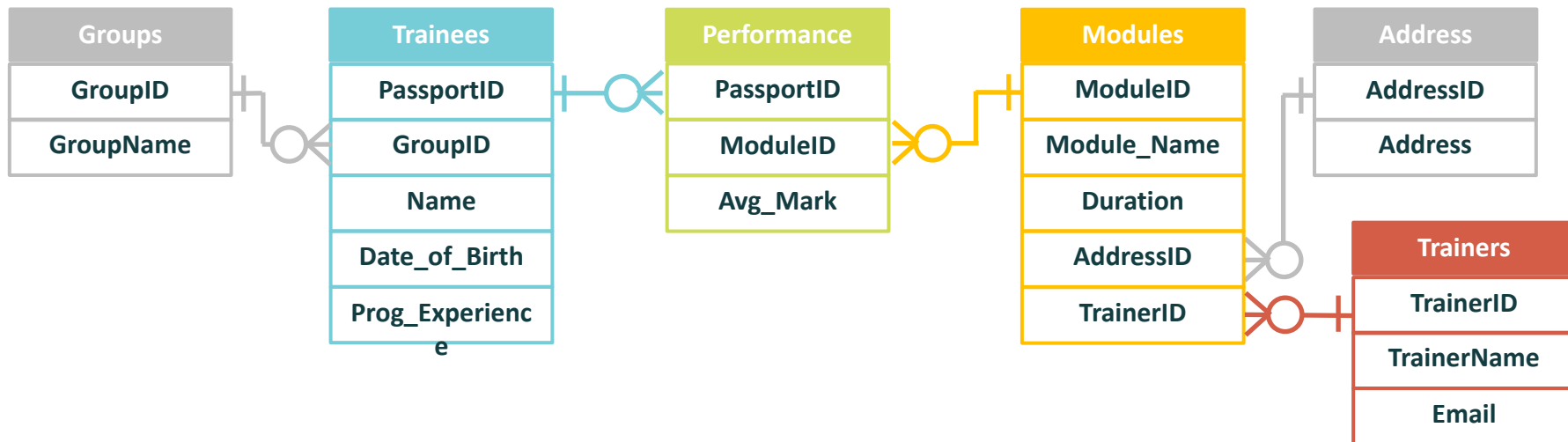
Conversion from 2NF to 3NF

Modules				
ModuleID	ModuleName	ModuleDuration	Address	AddressPhone
1	DB Basics	1	K1/1	+375299876543
2	SQL Foundation	4	DM3	+375293456789
3	SQL for Analysis	4	K1/1	+375299876543



Address	
Address	AddressPhone
K1/1	+375299876543
DM3	+375293456789

Disadvantages of Normalization

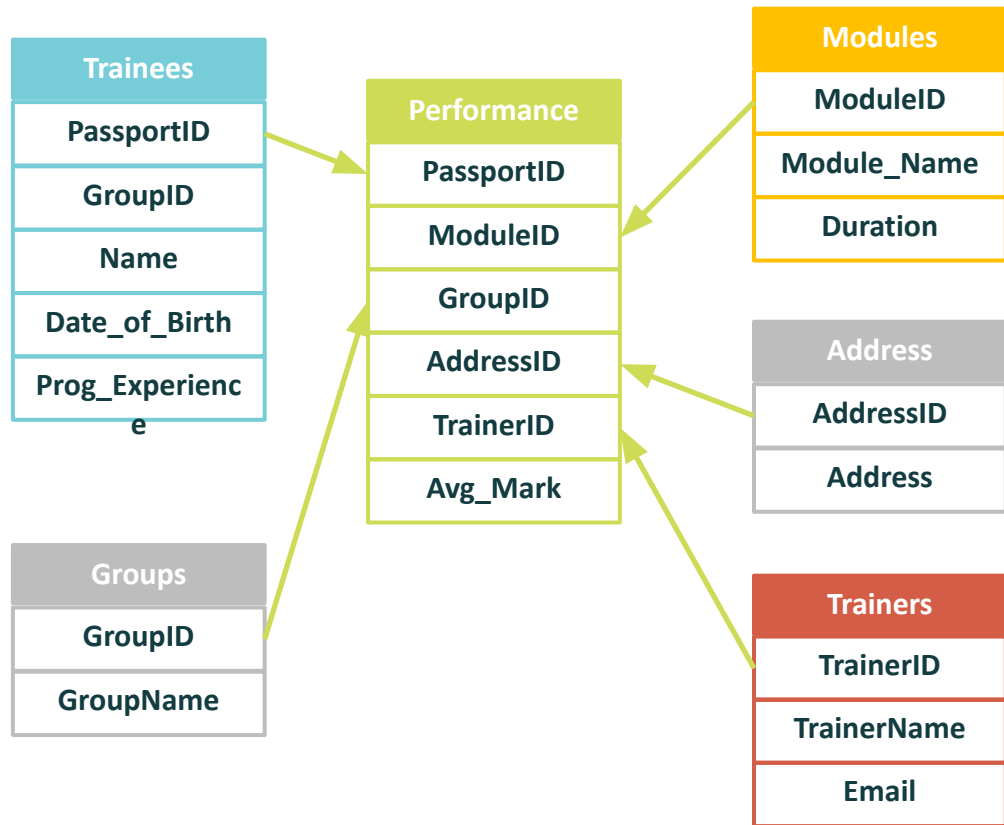


A completely normalized database needs clear and broad understanding of the business

More tables to join: by spreading out your data into more tables, you increase the need to join tables

The data model is optimized for applications, not for Ad-hoc querying

Denormalization



To enhance query performance

To make a database more convenient to manage

To facilitate and accelerate reporting

Does not mean not doing normalization.

It is an optimization technique that is applied **after** doing normalization

Data Warehousing

